DOCUMENT RESUME

ED 128 396 TM 005 550

AUTHOR Lolli, Anthony, Jr.; Owen, Steven V.
TITLE Student Ratings: What Is the Frame of Reference?
INSTITUTION Connecticut Univ., Storrs. Bureau of Educational Research.
PUB DATE [Apr 76]
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; Check Lists; College Students; College Teachers; Demography; Effective Teaching; *Evaluation Criteria; Expectation; Grade Point Average; Grade Prediction; *Higher Education; Multiple Regression Analysis; *Rating Scales; Student Characteristics; *Student Evaluation of Teacher Performance; Student Opinion; Teacher Behavior; Test Construction; Test Reliability; *Test Validity

ABSTRACT

Two issues in the controversy over the use, interpretation, validity, and relevance of student ratings are explored here: the definition of preferred teacher behavior in terms of the rating instrument used, and the relationship between the student's expected grade, grade point average, and the ratings recorded. The objectives of the present research were threefold: (1) to develop a short rating scale and associated diagnostic checklist; (2) to examine the construct validity and reliability of the scale; and (3) to assess the relationships between selected student demographic data and ratings on the scale. After a pilot test, there appeared to be little advantage to using an observed/preferred behavior format, so that format was discontinued in further revision of the scale. A repeated measures analysis of variance indicated that discrepant groups, i.e., those whose actual grade was two or more letter grades away from his/her expected grade, should be omitted from computations to increase the validity of student ratings. Further research was suggested to determine whether demographic variables interact with group discrepancy to influence evaluations of teacher behaviors. (Author/BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128396

BUREAU OF EDUCATIONAL RESEARCH AND SERVICE

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

IM005 550

RESEARCH REPORT SERIES

SCHOOL OF EDUCATION
THE UNIVERSITY OF CONNECTICUT

Student Ratings:

What is the Frame of Reference

Anthony Lolli, Jr. and Steven V. Owen

University of Connecticut

A paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Student Ratings:
What is the Frame of Reference

Anthony Lolli, Jr. and Steven V. Owen

University of Connecticut

Introduction.

Controversy over the use, intrepretation, validity, and relevance of student ratings continues. One particular issue that has received relatively little attention is the student's evaluative standard or frame of reference. Users of student ratings typically assume that students prefer the positive extreme in nearly all rating items. Yet students may not always prefer the extreme values in rating items. Not only is there variability of preference among students within a single class, there are also systematic shifts across classes in perceptions of the "ideal" teacher. These shifts appear to be strongly related to the student's major area of study (Slater and Owen, 1974).

Levinthal et al. (1971) asserted that student ratings of teacher behaviors imply comparative judgments between the values of the observer and his observations. Levinthal and his colleagues found that judgments of ideal or "preferred" behaviors varied across students and items, as did interactions between observed and preferred responses. These researchers called for additional research and use of a two scale (observed/preferred) rating format.

It seems possible that the meaning and validity of rating instruments would improve if students were asked both what they observed and what they would prefer to observe. For example, if the mean observed rating

on a five-point item were, say, 4.0, the instructor could apparently improve a full point on that item. However, if he discovered that the mean preferred rating on the same item were 4.0, there is not much room for improvement. Indeed, if a rating of 4.0 is the preferred locus, then an observed rating higher than 4.0 could not accurately be regarded as an improvement.

Another area concerned with the student's frame of reference was researched by Bausell and Magoon (1972). The question of whether a student's expected grade and GPA systematically affect subsequent teacher evaluations has not been answered clearly. The reason, in part, is that some studies have examined cases in which grades were awarded prior to the students' evaluation of their instructors, while other studies required students to evaluate their instructors prior to the awarding of grades. Bausell and Magoon found that when students were divided into groups on the basis of discrepancies between their GPA and their expected grade that there was a systematic effect on subsequent ratings. As might be predicted, those students with expected grades higher than their GPA appeared to give inflated ratings, while those students with low expected grades gave low ratings.

The objectives of the present research were threefold:

1) To develop a short rating scale and associated diagnostic checklist;

2) to examine the construct validity and reliability of the scale; and

3) to assess the relationships between selected student demographic data and ratings on the scale.

Pilot Study

A 10-item teacher evaluation instrument was constructed using five

point Likert questions pertaining to actual, observed teaching behaviors. In developing items, an effort was made to compromise between high inference statements (e.g., "attitude toward students") and low inference questions (e.g., "meets class on time"). Ten parallel items were constructed on which students indicated their preference for those behaviors (see Appendix A). Gagné and Allaire (cited in Crasha, 1975) suggested that use of a discrepancy format (difference between observed and "ideal" ratings) could help to account for student rating variation. One of the procedures conducted during this pilot tested this assertion in an incremental validity format. Observed item scores were regressed against two global, criterion items, resulting in multiple correlations of .86 and .70. Next, two multiple correlations were computed forcing the order of the predictor variables so that the observed ratings entered first, followed by difference scores (preferred items minus observed items) as a means of examining incremental validity. The resultant multiple correlations of .86 and .70 were identical to those regressions using the observed items only. Thus, there was no unique criterion variance accounted for by the difference scores. The assertion of Gagné and Allaire was not supported in this pilot study.

It was suggested, following intrepretations of the results of this pilot, that the observed/preferred format should not be employed for all questions since sufficiently large differences between the means of the observed and preferred items did not occur for all items. Lolli, Owen, and Froman (1975) suggested that, in determining which rating items qualify for an observed/preferred format, field tests should be undertaken for any new measures. As there appeared to be little advantage

to using an observed/preferred format with the current items, that format was discontinued in further revision of the scale.

Revised Instrument

Several of the pilot items were rewritten on the basis of difficulty in intrepretation, or because new items were believed to be more meaningful in terms of evaluative data. This revised instrument is shown in Appendix B. The demographic items appearing at the top of the instrument were included in order to assist in determining the validity of the instrument. Items 1 and 2 were designed to be global criterion items. An important criticism of teacher rating scales is the lack of external criterion measures which show that instructors who receive higher ratings do in fact produce higher amounts of learning in their students. Since external criterion measures are not immediately available, items 1 and 2 were included as the best available measure.

Items 3 through 13 were gleaned from existing rating instruments and from related literature. They represent issues which are suggested as important by recent research literature in the area of teacher effectiveness.

Items 14 through 64 are included for diagnostic purposes and are intended to provide the individual instructor with feedback related to specific behaviors which are particularly bothersome to students. It should be noted that items 14 through 64 are primarily the result of student input. Many of these items have appeared with high frequency in an open-ended "comments" section of a currently used rating scale. Notice that items 1 through 13 can be regarded as high inference measures, while, items 14 through 64 are more typical of low inference measures.

The high inference items should be regarded as "evaluative" measures (say, for promotion or tenure decisions), while the low inference items are meant to be used as feedback for the improvement of instruction.

Procedures.

Because of the many different procedures in this study, each separate description will be introduced with a brief research question that the procedure is intended to answer. The same format will be followed in the presentation of results. The scale was given to 4,930 graduate and undergraduate students during the last week of the 1975 fall semester. The students were drawn from 127 classes across 84 departments.

(What is the influence of expected grade, and GPA, on rating items?)

To examine the construct validity, a series of multivariate analyses were undertaken. Two stepwise multiple regressions were run. First, the 11 evaluative items were regressed against the students' expected grades. The purpose of this procedure was to determine the effect of expected grades on the rating items. Next, the 11 evaluative items were regressed against the students' grade point average (GPA) to determine the effect of GPA on the rating items.

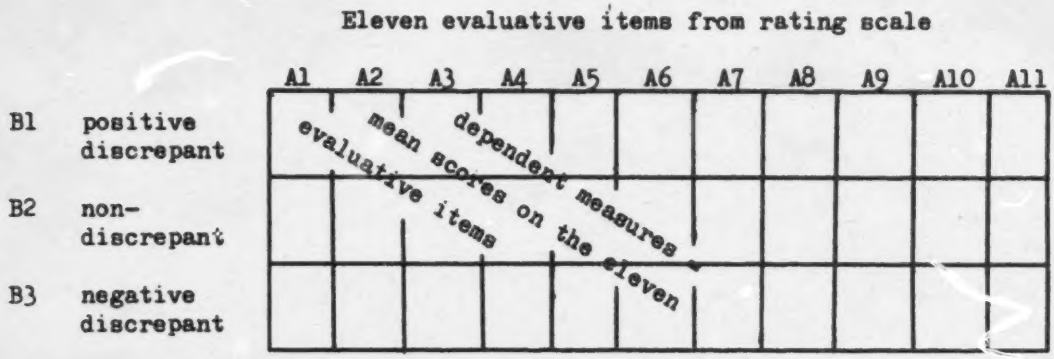
(Do discrepant and non-discrepant groups yield significantly different ratings?)

Third in this series of analyses, a Type I repeated measures ANOVA was run. The between-groups independent variable was student discrepant group membership (B main effect). The within-groups independent variables were the 11 evaluative items (A main effect). The dependent variables were the scores on the 11 evaluative items. This design is shown

in Figure 1, and follows the suggestion of Greenhouse and Geisser (1959) for interpreting profiles of data.

Figure 1

Type I Repeated Measures Analysis of Variance for Discrepant Student Groups and Evaluative Items



For the purpose of this analysis, a member of a discrepant group is any student whose QPR is two or more letter grades away from his/her expected grade. In other words, a student belongs to a discrepant group if that student:

1) Is an "A" student but expects a C, D, or F;
2) is a "B" student but expects a D or F;
3) is a "C" student but expects an A or F; or
4) is a "D" student but expects an A or B.

The discrepant group was divided into two types:

1) Students whose expected grade is two or more letter grades above their QPR ("positive" discrepant group), and

2) students whose expected grade is two or more letter grades below their QPR ("negative" discrepant group).

(What are the psychological dimensions underlying the evaluative portion of this instrument?)

Next, a principal component analysis was performed on the evaluative items. Orthogonal and oblique rotations followed the component analysis. Alpha internal consistency estimates were calculated for the meaningful factors.

(What is the influence of criterion item 2, and demographic items, on rating items?)

Two stepwise multiple regressions were run against criterion item 2 to test the adequacy of the instrument. First, the 11 evaluative items were regressed against criterion item 2 to determine the extent that the overall teacher rating question could be predicted from the 11 evaluative items. Next, the demographic items (see Appendix B) were regressed against criterion item 2. The purpose here was to determine whether the several demographic items were related to the global criterion item 2.

A preliminary analysis examined the correlation between criterion item 1 (How much have you learned in this course?) and subsequent actual mid-term exam scores for a sample of 169 students in an undergraduate educational psychology course. The correlation was .11. As the internal consistency estimate of the exam was .89, and the equivalence reliability estimate was .81, it was clear that student estimates of "amount learned" did not overlap as much as expected with exam scores. Obviously, amount

learned does not take into account the entering behavior (or knowledge) of the students. Perhaps a more accurate criterion would have been the residual gain between entry behavior and exit behavior. In any case, it was decided that item 1 was not a valid criterion measure and it was dropped from further analyses.

Results

Table 1 presents the means and standard deviations for the 11 evaluative items.

Insert Table 1

(What is the effect of expected grade, GPA, criterion item 2, and demographic items, on rating items?)

Table 2 summarizes the regression analyses and lists the dependent variables, independent variables, multiple correlation coefficients, and "general outcomes" (the extent to which the dependent variable can be predicted from the independent variables) for each multiple regression.

Insert Table 2

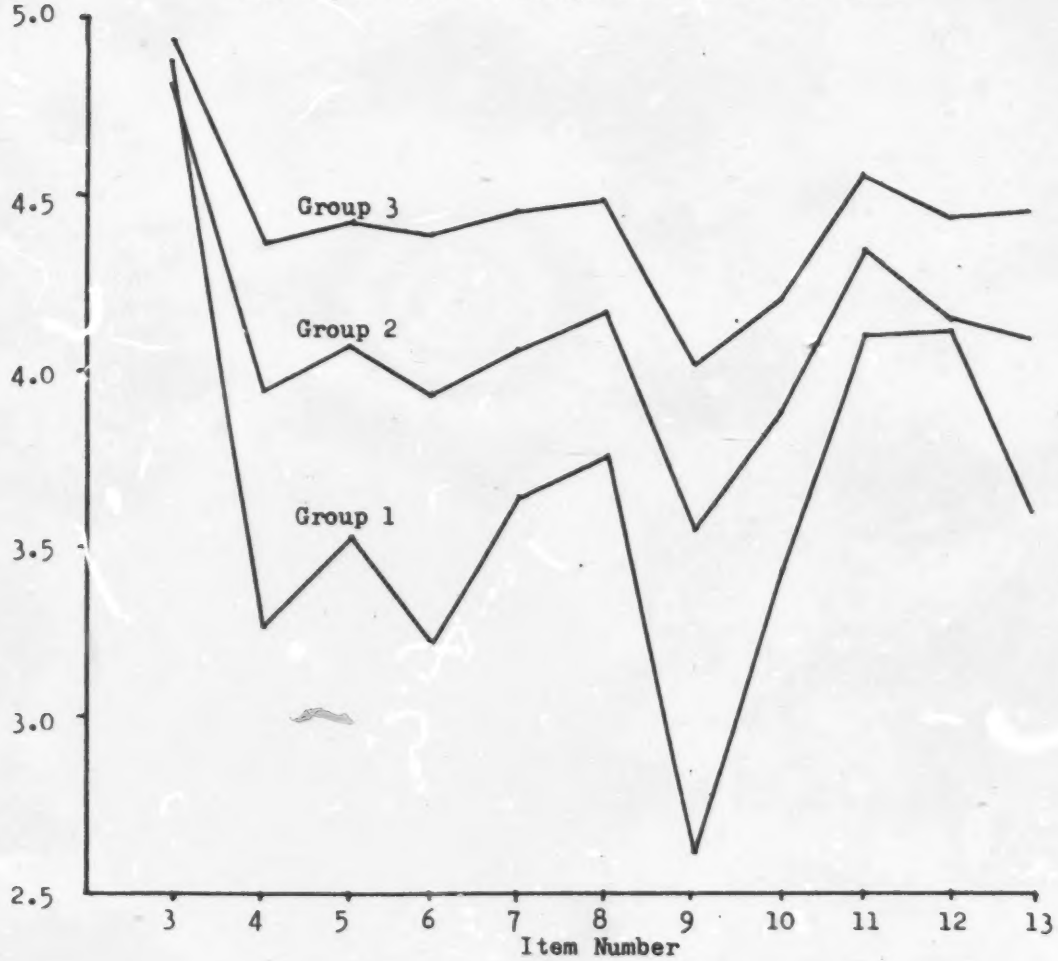
(Do discrepant and non-discrepant groups yield significantly different ratings?)

The results of the repeated measures ANOVA are shown in Table 3. It should be noted that the B main effect was significant beyond the .001 level. Means for each group across each item are shown in Table 4. As a result of the significant A x B interaction the group means across the

11 evaluative items are plotted in Figure 2. With the exception of the first evaluative item all interactions were ordinal. It will be shown in the report of the results of the factor analysis that this item (evaluative item 1) was a "splinter" item and did not load on either of the meaningful factors.

Figure 2

Type I Repeated Measures Analysis of Variance for Discrepant Student Groups and Evaluative Items



Insert Table 3

Because of the ever present problem of missing data, four exploratory principal component analyses were run: two employing pairwise deletion and two using casewise deletion of data for both orthogonal and oblique solutions. The method of pairwise deletion was selected over casewise deletion because both options yielded highly similar loadings and factor structures for each type of rotation, and because the pairwise deletion option included more information in its analysis.[1] The oblique rotation was ultimately selected because it yielded a cleaner analysis with less overlap of factors, and made better psychological sense. In addition, the primary (unrotated) axes were somewhat correlated with each other ($r = .26$).

Table 4 contains the primary pattern loading matrix, oblique solution for the eleven evaluative items, and alpha internal consistency estimates. Only items with loadings of .30 or higher are included. Using a root criterion of unity, two dimensions were generated and both were meaningful enough to be retained. The two component solution accounted for 58.1 percent of the total variance associated with item interrelationships. The items of unit complexity (loading highly on only one factor) were given principal consideration in naming the factors. The content of the actual item stems was used to generate the factor descriptions.

---

[1] Casewise deletion removes all information from a subject if he is missing only a single bit of data. Pairwise deletion, by contrast, removes only the particular bit of information missing. For a discussion of benefits and limitations of each type of deletion, see Nie et al., 1975, pp. 353; 503-4.

Insert Table 4

Factor I (47.5 percent of the total variance) was named Teaching Effectiveness. The item content defining the factor suggests that instructors receiving a high score on this factor were perceived by their students as displaying behaviors within the classroom which help them to make clear explanations of their well organized materials and procedures. Factor II (10.6 percent of the total variance) was named Interaction With Students, and reflects the quality of exchange that takes place between the instructor and his/her students considered apart from the quality of the instructor's teaching behaviors per se.

Discussion.

The results of the reliabilities of both dimensions which emerged from the factor analysis were reported in Table 4. The coefficients are adequate to warrant their use in examining student ratings of teacher behavior through the implementation of this instrument. Good intentions to factor the 51 diagnostic items were thwarted when a $98.00 computer run depleted our account. (The factor analysis was begun, but was aborted before rotations were performed.) This analysis will be completed shortly, and is meant to be a confirmatory factor analysis. That is, it will test hypotheses about which diagnostic items "fit" with which evaluative, high inference items. It is hoped that the diagnostic items will generate comparable factor structures that will allow instructors being rated to obtain additional information from the diagnostic items that will be able to

further explain poor ratings on given evaluative items. In this way, the instructor who wishes to can use this instrument as a basis for improving teaching effectiveness by targeting specific behaviors which are troublesome with respect to the frame of reference of the consumers (i.e., students).

One of the more frequent criticisms of student ratings of teacher behaviors is that students who do well in a course will give the instructor a high rating, while students who do poorly will rate the instructor lower. A similar, parallel argument might be made for students with high versus low GPAs. The respective multiple correlations of .25 and .07 for the dependent variables expected grade and GPA (predicted by the 11 evaluative items) provide only weak support for that criticism. However, when both expected grade and QPR are combined (discrepant groups) we find an important, different conclusion. The group means, collapsing across the 11 items are significantly different. In addition, the direction of the bias conforms to the criticism stated above. The latter outcomes can be seen in the repeated measures analysis; in particular, refer to Figure 2.

A probable explanation for the contradictory findings yielded from the two multiple regressions versus the results from the repeated measures ANOVA is that since the non-discrepant group is so much larger than either of the discrepant groups that in both multiple regressions the total effect is to "wash out" the biasing effect of the discrepant groups' ratings. This biasing effect can only be seen clearly when the groups are compared through the repeated measures ANOVA.

It seems clear that the discrepancy between QPR and expected grade should be considered by investigators studying student ratings of teacher

behaviors. On the basis of these results, we suggest that the evaluations by students who fall into either of the discrepant groups be deleted from the subsequent calculations. This will probably result in a more accurate picture of the resultant ratings since the validity of the student ratings as a group is increased. This proposed method of data reduction is especially important for those classes which have small numbers of students, since the biasing effect of discrepant evaluations could be increased if there is a disproportionate ratio of discrepant students in that class. Such a ratio would result in questionable evaluations of that particular instructor. One example that comes to mind is the course that is required, and taught for non-majors. It is not unreasonable to expect that the students who find themselves faced with a new and unfamiliar body of knowledge might expect to do less well than their QPR would indicate.

The results of the stepwise multiple regressions in which the demographic items were regressed against criterion item 2 yielded a multiple correlation of .25 (low predictability). However, in light of the findings stated above, the relationship of demographic variables to student ratings is an area which requires additional research. For instance, the interaction between the strongest demographic variable and group discrepancy needs to be tested.

Summary.

A rating scale employing Likert-type questions, two criterion questions, demographic items, and diagnostic items was constructed. The observed/preferred format, as a means of examining the student's frame

of reference appears to be of questionable value, at least for the evaluative items appearing on this scale. Principal component analyses yielded two factors (Teaching Effectiveness and Interaction with Students). Resulting Alpha internal consistency coefficients were adequate to warrant the use of the instrument. Evidence was presented which suggested that while individual multiple regressions of evaluative items against 1) QPR, and 2) expected grade demonstrated no substantial effect. The results of a repeated measures ANOVA demonstrated a biasing effect on ratings by discrepant groups. It was suggested that in order to increase the validity of student ratings, the evaluations made by discrepant group members should be omitted from computations. Further research was suggested to determine whether demographic variables interact with group discrepancy to influence evaluations of teacher behaviors.

References

Bausell, R.B. & Magoon, J. Expected grade in a course, grade point average, and student ratings of the course and the instructor. Educ. & Psychol. Meas., 1972, 32, 1013-1023.

Brown, D.L. Faculty ratings and student grades: a university-wide multiple regression analysis. J. Educ. Psych., in press.

Grasha, A. The role of internal instructor frames of reference in the student rating process. J. Educ. Psych., 1975,

Greenhouse, S.W. & Geisser, S. On methods in the analysis of profile data. Psychometrika, 24(2), 95-112.

Levinthal, C., Lansky, L., & Andrews, O. Student evaluation of teacher behaviors as estimations of real-ideal discrepancies: a critique of teacher rating methods. J. Educ. Psych., 1971, 62, 104-109.

Lolli, A., Owen, S.V., and Froman, R.D. Student ratings: further explorations in validity. Paper presented at the Northeastern Educational Research Association, Ellenville, N.Y., October 1975.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. SPSS: Statistical package for the social sciences (2nd ed.). N.Y.: McGraw Hill, 1975.

Slater, J.K. & Owen, S.V. Departmental differences in student perception of the "ideal" professor. Paper presented at the Northeastern Educational Research Association, Ellenville, N.Y., October 1974.

Table 1

Item Means and Standard Deviations for Evaluative Items Across Three Discrepant Groups

Item Number	Item Stem	Mean				Standard Deviation			
		Group 1 (N = 28)	Group 2 (N = 3546)	Group 3 (N = 364)	Total	Group 1	Group 2	Group 3	Total
3	Meets class regularly	4.89	4.83	4.93	4.84	0.31	0.52	0.29	0.50
4	Presents material clearly	3.30	3.91	4.25	3.93	1.38	1.05	0.86	1.04
5	Demonstrates organization	3.50	4.04	4.33	4.07	1.24	0.99	0.77	0.98
6	Makes purpose clear	3.29	3.91	4.22	3.94	1.33	1.08	0.89	1.07
7	Fulfills objectives	3.57	4.06	4.34	4.08	1.27	0.94	0.78	0.93
8	Makes assignments clear	3.86	4.18	4.40	4.20	1.19	1.02	0.89	1.01
9	Stimulates interest	2.64	3.50	3.92	3.54	1.45	1.13	1.00	1.18
10	Important exam items	3.37	3.76	4.15	3.79	1.37	1.13	0.95	1.12
11	Grades fairly	4.11	4.31	4.57	4.33	1.17	0.92	0.73	0.91
12	Is accessable	4.19	4.15	4.39	4.17	0.77	0.95	0.76	0.94
13	Shows interest in students	3.70	4.08	4.36	4.10	1.24	1.07	0.89	1.06

Table 2

Summary of Stepwise Regression Analyses

Dependent Variable	Independent Variables	R	General Outcome
Expected Grade	Evaluative Items	.25	Low Predictability
QPR	Evaluative Items	.07	No Relationship
Criterion 2 (rate the instructor)	Evaluative Items	.82	Strong Relationship
Criterion 2 (rate the instructor)	Demographic Items	.25	Low Predictability

Table 3

Source Table for Repeated Measures ANOVA with Three Student Groups Across Evaluative Items

Source	SS	df	MS	F
Between Ss				
B (discrepancy groups)	186.54	2	93.27	18.99***
error	14606.28	2974	4.91	
Within Ss				
A (eleven items)	289.39	10	28.94	52.35***
AB	51.46	20	2.57	4.65***
error	16440.36	29740	0.55	

*** $p < .001$.

Table 4

Factor Loading Matrix for Evaluative Items Derived From Principle Component Analysis and Oblique Solution[1]

Item Number	Item Stem	Loading
Factor I:	Teaching Effectiveness (Alpha estimate = .78)	
4	Presents material clearly and effectively	.86
5	Demonstrates overall organization	.86
6	Make course objectives clear	.84
7	Fulfills class objectives	.82
8	Makes assignments clear	.53
9	Stimulates interest	.64
10	Exam items stress important aspects	.46
Factor II:	Interaction with Students (Alpha estimate = .65)	
11	Grades fairly and impartially	.39
12	Is accessible to students	.79
13	Shows interest in and concern for students	.76

[1]The entire matrix is available from the senior author; this table only shows items which loaded .30 or higher.

Appendix A

The University of Connecticut Survey of Courses and Teaching

Demographic Information

Your semester:[ ][ ][ ][ ][ ][ ][ ][ ][ ]Is this course in your major field?[yes][no]
1 2 3 4 5 6 7 8 Grad

Expected grade in this course:[ ][ ][ ][ ][ ][ ]
A B C D F Pass

Cumulative grade average:
1.0-1.5[ ] 1.5-2.0[ ] 2.0-2.5[ ]
2.5-3.0[ ] 3.0-3.5[ ] 3.5-4.0[ ]

Percent of class attendence: 0-24%[ ] 25-49%[ ]
50-74%[ ] 75-89%[ ] 90-100%[ ]

Time spent on this course compared to other courses of equal credit:
considerably more[ ] somewhat more[ ] about the same[ ]
somewhat less[ ] considerably less[ ]

Observed/Preferred Items

For the following ten questions, rate each item in two ways:
a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you're satisfied,the 2nd scale will be marked the same as the first.

From my observations, this instructor:	almost never	not very often	sometimes	most of the time	almost always
1. presents material in a clear & effective manner					
(preferred)					
2. stimulates interest					
(preferred)					
3. makes work assignments & student responsibilities clear					
(preferred)					
4. uses exam items which stress important aspects of the course					
(preferred)					
5. grades fairly and impartially					
(preferred)					
6. demonstrates overall organization					
(preferred)					
7. meets class regularly and on time					
(preferred)					
8. fulfills class objectives and obligations					
(preferred)					
9. is accessible to students both in and out of class					
(preferred)					
10. shows an interest in and concern for students					
(preferred)					

Global Criterion Items

11. How much have you learned from this course?
almost nothing[ ] a little[ ] moderate amount[ ] a great deal[ ]

12. How would you rate this instructor in general?
poor[ ] less than adequate[ ] average[ ] good[ ] excellent[ ]

Appendix B

The University of Connecticut Survey of Courses and Teaching

Department ____ Course # ____ Section ____ Branch ____ Instructor ____

DIRECTIONS: PLEASE USE A NO. 2 PENCIL TO FILL IN APPROPRIATE BLOCKS. DO NOT MAKE STRAY MARKS. DO NOT SIGN YOUR NAME. ERASE CLEANLY IF YOU CHANGE AN ANSWER. RESULTS WILL NOT BE SEEN BY THE INSTRUCTOR BEFORE THE COURSE IS COMPLETED.

Semester Standing	Is Course in Your Major	Expected Grade in This Course	Cumulative Average (QPR)	How Often Did you Attend Class
1-2	Yes	A	Less than 10	90-100%
3-4	No	B	10-14	75-89%
5-6		C	15-19	50-74%
7-8		D	20-24	25-49%
9 or more		F	25-29	0-24%
Graduate		Pass	30-34	
		Audit	35 or more	

1. How much have you learned from this course? ☐ almost nothing ☐ a little ☐ a moderate amount ☐ quite a bit ☐ a great deal
2. How would you rate this instructor in general all around teaching ability? ☐ poor ☐ less than adequate ☐ average ☐ good ☐ excellent

Please omit those of the following items which are not appropriate

In my opinion, this instructor

	almost never	occasionally	sometimes	most of the time	almost always
3. Meets class regularly and on time	☐	☐	☐	☐	☐
4. Presents material in a clear and effective manner	☐	☐	☐	☐	☐
5. Demonstrates overall organization	☐	☐	☐	☐	☐
6. Makes purpose and objectives of course clear	☐	☐	☐	☐	☐
7. Fulfills class objectives	☐	☐	☐	☐	☐
8. Makes work assignments and student responsibility clear	☐	☐	☐	☐	☐
9. Stimulates interest	☐	☐	☐	☐	☐
10. Uses examination items which stress important aspects of course	☐	☐	☐	☐	☐
11. Grades fairly and impartially	☐	☐	☐	☐	☐
12. Is accessible to students both in and out of class	☐	☐	☐	☐	☐
13. Shows an interest in and concern for students	☐	☐	☐	☐	☐

DIAGNOSTIC CHECK LIST: The following is for the use of your instructor for the purpose of identifying areas of performance which need improvement. Please check ONLY those areas which you found to be PARTICULARLY TROUBLESOME OR BOTHERSOME.

THE INSTRUCTOR:

14. ☐ was consistently late in starting or ending class
15. ☐ was consistently unprepared
16. ☐ had illegible handwriting
17. ☐ came to class intoxicated or otherwise incapacitated
18. ☐ had a distracting smoking habit
19. ☐ seemed unfriendly toward students
20. ☐ criticised and embarrassed students unfairly
21. ☐ was patronizing, talked down to students
22. ☐ maintained attitude of "I am always right"
23. ☐ enforced pointless rules
24. ☐ seemed preoccupied with personal problems
25. ☐ neglected course for other activities
26. ☐ was too dogmatic or opinionated to present materials fairly
27. ☐ displayed favoritism toward certain students
28. ☐ displayed prejudice on the basis of race, sex, religion, ethnicity, etc.
28. ☐ spoke in a monotone
30. ☐ was difficult to hear or understand (mumbles, accent, inaudible)
31. ☐ had distracting speech habits
32. ☐ had distracting mannerisms (nervousness, pacing, twitch, etc.)
33. ☐ used too much profanity, vulgarity
34. ☐ engaged in too much idle chit-chat
35. ☐ just read lectures
36. ☐ presented dull, boring lectures
37. ☐ did not cover materials in depth, too superficial
38. ☐ presented too much detail or trivia
39. ☐ did not stress important points
40. ☐ presented material at too fast a pace
41. ☐ presented material at too slow a pace
42. ☐ went over the same thing too often
43. ☐ did not use enough illustrations, examples
44. ☐ spent too much time on class discussions
45. ☐ spent too little time on class discussions
46. ☐ responded to questions in a vague and confusing manner
47. ☐ assigned reading materials and then ignored them
48. ☐ did not explain what was expected on exams or assignments
49. ☐ did not provide enough feedback or comments on exams or papers
50. ☐ was slow in returning graded work
51. ☐ stressed rote memorization or trivia rather than understanding on exams
52. ☐ applied grading standards inconsistently
53. ☐ grades too easily
54. ☐ did not provide enough exams or other materials for evaluation
55. ☐ required purchase of costly materials which were not worth the price
56. ☐ assigned too much "busy work"
57. ☐ makes unreasonably long and burdensome assignments

THIS COURSE:

58. ☐ too much smoking took place
59. ☐ cheating was widespread
60. ☐ enrollments were too large for effective instruction
61. ☐ content was not sufficiently challenging
62. ☐ content was over my head, too difficult
63. ☐ content did not match expectations, catalog description
64. ☐ too much responsibility was delegated to graduate assistant